THE SEMANTIC WEB AND TURNS OF SOCIAL SCIENCE

Complete Research

Kosaka, Takeshi, Tokyo University of Science, Suwa, Japan, kosaka@rs.tus.ac.jp

Abstract

This research is a theoretical investigation of the semantic Web in light of the turns of social science. Social science has experienced several turns to have better understandings of social realities. If it is consistent with the turns, then the semantic Web is expected to have better representation of social realities than the conventional information systems. There are two major languages in the semantic Web: RDF and OWL. It is revealed that describing concepts by RDF conforms to the idea of phenomenology. Since RDF and OWL are based on the open-world assumption, the semantic Web is considered to be consistent with the linguistic turn. With the high consistency with social science, the semantic Web can be a good candidate of IS research topics. IS research is expected to contribute to the development of the semantic web, and the development of methodology in particular.

Keywords: The semantic Web, Social science, Open-world assumption, IS discipline.

1 Introduction

This research is a theoretical investigation of the semantic Web in light of the turns of social science. Social science has experienced several turns to have better understandings of social realities. If the semantic Web is consistent with the turns, then the semantic Web is expected to have better representation of social realities.

What is the semantic Web? There is increasing awareness of Linked Open Data (LOD) as a new sharing method of data separately prepared in various places such as libraries, museums, etc. LOD is the realization of the semantic Web which promotes the interoperability among data independently constructed by making the meanings of data machine-processable. With these current moves, the term “the semantic Web” has been exposed more than ever.

The brief definition of the semantic Web follows. The semantic Web describes the relationships between things (like A is a part of B and Y is a member of Z) and the properties of things (like size, weight, age, and price). Berners-Le (1999) gives us a lucid explanation: “If HTML and the Web made all the online documents look like one huge book, semantic Web technologies (RDF, schema, and inference languages) will make all the data in the world look like one huge database.” He and his colleagues (2001) say further: “The semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation.”

The semantic Web technology enables humans to create data on the web, to build vocabulary, and to write rules for handling the data. The semantic Web technology consists of OWL and RDF languages and others. The design of RDF is intended to meet the following goals: having a simple data model, having formal semantics and provable inference, allowing anyone to make statements about any resource and others (W3C RDF 2004). That is to say, the semantic Web is the distributed knowledge and data in a re-usable format.
The main reason why we discuss the semantic Web is a relation to information systems (IS). IS development is widely known as a failure-prone project through the Standish Chaos reports and various articles (Glass, 2006). The major reasons of failure are that developers face difficulty in capturing user needs because they are not pre-existing (Boland, 1991), and that there is a contrasting structure of an inflexible IS and the changing environments (Truex et al, 1999). We discuss the semantic Web as an approach that brings a possible solution to these problems.

It is thought that IS failure comes from the characteristics of database residing in as a core of IS. The existing database of relational DB model is based on the closed-world assumption (CWA), while the Semantic Web is based on the open-world assumption (OWA). It is thought that the OWA is better suited with the distributed development and the adaptation to the changing world than the CWA.

There are two sub reasons or objectives. First, we have to refer to an argument (Weber, 1999) that the articulation of a foundation is essential if the information systems discipline is to prosper both intellectually and pragmatically. Benbasat and Zumud (2003) discuss why establishing an identity for the IS field is important. They argue that a dominant design or a foundation for the IS discipline has yet to be realized. The areas unique to IS discipline are considered to be what could not be explained adequately by theories from other disciplines. According to Weber (1999), they are conceptual modelling, semantic modelling, and data modelling. Since the semantic Web is concerned with semantic modelling, we think that the semantic Web is worth discussing in relation to IS discipline.

Secondly, it is thought that not only needs of organizations but also those of individual persons become important in distributed development environments. The age is coming, that is, the age that an individual person writes data, and requires and uses data. With this move in distributed environments, systems analysis (SA) methodology is expected to become necessary for capturing personal or subjective needs in addition to organizational or objective needs. We are making efforts in building the SA method for the personal use (Kosaka, 2013). This theoretical investigation of the semantic Web is expected to reveal and justify the development needs of this new SA methodology.

This research discusses the feasibility of the semantic Web as a possible research topic in the IS research field by examining whether the semantic Web has a consistency with the turns of social science. Social science not only depends on the traditional positivistic approach but also has developed its own epistemology to have better understanding of societies and individuals. If it has the higher consistency with the turns of social science, it is thought that the semantic Web is an appropriate research area or topic in the IS research.

2 The semantic Web as an object of research

2.1 The layer cake of the semantic Web

The object of this research is the semantic web. Its architecture is illustrated by the semantic Web layer cake as shown in Fig.1. The main languages in the layer cake are RDF and OWL. They are accepted as endorsed languages by W3C. IRI (International Resource Identifier) that is a generalization of URI is used as an identifier. XML syntax is used for all description.

RDF (Resource Definition Framework) is a framework used to produce a statement or data in a so-called triple. RDFS (RDF schema) provides a data-modelling vocabulary for RDF statements. OWL (Web Ontology Language) adds more high-level constructs to RDFS to describe the meaning of RDF statements. For example, OWL is used to express parent-child relations and inclusion relations among classes. SPARQL is a query language of RDF statements. Proof and Trust are not yet realized, and are considered inter-subjective processing.

An RDF statement expresses a relationship between two resources by the following structure: <subject><predicate><object>. The subject and the object represent the two resources being related by
the predicate, which represents the nature of their relationship. The relationship is phrased in a directional way from the subject to the object, and is called a property in RDF. Because RDF statements consist of three elements they are called triples.

Here we show an example of RDF triples (W3C RDF1.1 Primer, 2014) informally expressed, although RDF statements are expressed formally using XML and IRI.

<Bob> <is a> <person>.
<Bob> <is a friend of> <Alice>.
<Bob> <is born on> <the 4th of July 1990>.
<Bob> <is interested in> <the Mona Lisa>.
<Bob> <has full name of> <Robert Smith>.
<the Mona Lisa> <was created by> <Leonardo da Vinci>.
<the video 'La Joconde à Washington'> <is about> <the Mona Lisa>.

Triples can be visualized as a connected graph. Graphs consist of nodes and arcs. The subjects and objects make up the nodes in the graph. The predicates form the arcs. The same resource can be a subject in a triple, and an object in another triple. Fig. 2 shows the graph of the sample triples. Once we have RDF statements like this we can use SPARQL to query for e.g. people interested in Mona Lisa.

2.2 Open-world and non-unique name assumptions

We are often encountered with a term “open-world assumption” (OWA) when discussing the semantic Web. W3C OWL-Guide (2004) mentions that the semantic Web is based on OWA as follows. “RDF and OWL Full are designed for systems in which data may be widely distributed (e.g., the Web). As such a system becomes larger, it becomes both impractical and virtually impossible to know where all of the data in the system are located. Therefore, one cannot generally assume that data obtained from such a system are complete. If some data appear to be missing, one has to assume, in general, that the data might exist somewhere else in the system. This assumption, roughly speaking, is known as the open world assumption.”

W3C OWL2 Premier (2009) mentions closed-world assumption (CWA) as follows. “If some fact is not present in a database, it is usually considered false (the so-called closed-world assumption) whereas in the case of an OWL 2 document it may simply be missing (but possibly true), following the open-world assumption.”

In the semantic Web, we assume that “our knowledge of the world might be incomplete and therefore any information not explicitly specified (or such that cannot be derived from the known data) is considered unknown” (Levin, 2008). This line of reasoning is called the monotonic reasoning. For further information of OWA and
the monotonic reasoning, see Patel-Schneider et al (2006).

OWA is considered an appropriate assumption in the world that is gradually constructed by multiple participants who develop information or knowledge in a distributed fashion. However, there is another case such as a query of flight information. The query is performed in the non-monotonic reasoning such that non-existence of information means false. An assumption that data expressed in a database is all is known as CWA, in which queries use Negation by Failure (NAF). In some cases the composite of OWA and CWA might be useful.

The semantic Web is also based on the non-unique name assumption. In the semantic Web where multiple participants create information independently from each other, it can be anticipated that different names are used for the same object, and vice versa. Therefore, the semantic Web employs the non-unique name assumption. It is possible to explicitly declare at the OWL level that different names show the same object and the same names show the different objects. The former case is expressed by owl.sameAs, and the latter by owl.differentFrom.

Concerning the difference between the semantic Web and the existing DB, we can refer to a research paper (Patel-Schneider et al, 2006) which studied the general distinction between the classical paradigm based on OWA and the datalog paradigm on CWA. The semantic Web belongs to the classical paradigm, and RDB (relational database) to the datalog paradigm.

Bergman (2009) shows the influences of this distinction between OWA and CWA in IS development brought by the use of the semantic Web and RDB. He argues that IS which uses RDB enforcing a single rigid schema, i.e., a closed world, tend to structurally fail in the age of changing environments and of the increasing importance of interconnectivity.

2.3 The open-world assumption and the layer cake

There is an idea to realize interoperability among data developed locally in distributed environment, the semantic Web. The layer cake has a stack of high expressive language in the upper layers. However, it is still not clear how the upper layers, i.e., proof and trust layers, are going to be implemented.

As mentioned later, we are thinking that they are realized not technically but socially. The top layer, trust, is expected to be of social realization in particular. It means that there must be a mechanism that brings reliability to a conclusion derived from knowledge or data in distributed environment. The top layer, trust, is expected to be gained by wide participation and use of many people. The proof layer is expected to identify people or institution whose knowledge the derived conclusion comes from.

3 Literature review

Research of the semantic Web in the viewpoint of social science is very restrictive. Therefore, we expand the range of literature. Firstly, there are many applications of the semantic Web. For example, we can find applications of RDF in EUROPEANA as a portal of libraries in Europe, DATA.GOV in open government in USA and so on. For an AIS conference paper, Jayadianti et al (2012) report a case in which they enabled data interoperability by building a common ontology with the use of RDF and OWL where each local government has developed data in different formats. From these examples we can understand that a part of the semantic Web, RDF in particular, has gained some social acceptance.

For a technological problem, there are researches showing that humans are good at combining concepts, however, the semantic Web cannot be expected to be so. For instance, humans often categorize objects according to the similarity between the objects, however, similarity is not a notion that can be expressed in a natural way in a web ontology language (Gärdenfors, 2004). Shirky (2003) argues that the semantic Web is just a machine for creating syllogisms. This issue is not considered a
problem in our research, because the semantic Web is a web of data but is not a replacement of human complex thinking.

The semantic Web includes not only RDF but also OWL for ontologies. If we expand prior literature up to ontology, then we can find several research papers showing the lack of ontology construction methodology. For example, there is no single correct ontology for any domain. Ontology design is a creative process and no two ontologies designed by different people would be the same (Noy et al, 2001). Ontology building is still more of a craft than an engineering task (Pinto et al, 2004). Ontologies often need to be built in a decentralized way, ontologies must be given to a community in a way such that individuals have partial autonomy over them, ontologies have a life cycle that involves an iteration back and forth between construction/ modification and use and ontologies should support the participation of non-expert users in ontology engineering processes (Pinto et al, 2009). Support for their participation is mostly lacking in these methodologies (ibid). The results achieved by the ontology engineering community in the last decades are of incontestable value for the large-scale uptake of semantic technologies (Simperl et al, 2010). However, similarly to other engineering disciplines, requirements analysis remains a challenging task (ibid). The participation of the model user in model construction is essential…This is because a constructivist understanding of conceptual modeling assumes different worldviews of the stakeholders …(Esswein et al, 2013). These discussions about the needs of methodology present the development needs for the semantic Web methodology to IS research as a social science. That is, these suggest indirectly the necessity to investigate the consistency of the semantic Web to social science prior to the development of methodology.

Research with direct relations to social science is very restrictive. Kroeeze (2010) is exceptional, and he mentions that IS development is grounded in a typical modernist view of static information and singular meaning, while formal ontologies – and the plural form of the word – may be regarded as a typical effect of postmodern trends in IS. He further says that creating ontology does not only require technical skills, but also analytical abilities and philosophical wisdom.

As reviewed above, prior literature of the semantic Web directly related to social science is so scarce that we are sure there is a significance or good reason to continue our research.

4 Turns of social science as a research framework

Social science has experienced several turns to have better understandings of social realities. If the semantic Web is consistent with the turns, then the semantic Web is expected to have better representation of social realities. Prior to the discussion of consistency, we have a brief introduction of the turns of social science that we use as a research framework.

Positivism has long dominated various fields. Husserl (1970, pp.5-6) also pointed out like this: The total world-view of modern man, in the second half of the 19th century, let itself be determined by the positive sciences and be blinded by the ‘prosperity’ they produced. During the 20th century, however, philosophy experienced several turns. The first as these was the phenomenological turn initiated by authors such as Husserl and Heidegger (Monod et al, 2006).

4.1 The phenomenological turn

Merleau-Ponty (2002, pp.xxii-xxiii) mentions: “The phenomenological world is not the bringing to explicit expression of a pre-existing being, but the laying down of being.” In phenomenology, we firstly make phenomenological reduction, i.e., returning to consciousness, then we constitute or re-constitute the world from there. It is a conscious activity that we see how we constitute or construct our world and various objects in our world.
4.2 The linguistic turn

It is in the early 20th century that the new movement occurred to rethink the working of language. One of them is the movement associated with Saussure (1857-1913), and the other is with Wittgenstein (1889-1951).

Until Saussure and Pierce came to the scene, languages are regarded as transparent media, and was not associated with any problems. In those days, Saussure separated a correspondence between reality and word (reality - word), and inserted a concept or idea in between them (reality, concept – sound). ‘Signified’ for concept and ‘signifier’ for sound constitute a sign. A system of signs segments and articulates the world as such and lays down the world as a meaningful existence. Saussure (1993, p.398) explains this situation like this: “The different ideas represent nothing pre-existing. There are no: a) ideas already established and quite distinct from one another, b) signs for these ideas. But there is nothing at all distinct in thought before the linguistic sign.” This thinking turned the way we see language 180 degrees.

Wittgenstein (1958, §.1) also mentions, “it seems to me, give us a particular picture of the essence of human language. It is this: the individual words in language name objects—sentences are combinations of such names.” Wittgenstein discarded this thinking of language, revealed “the limits of my language mean the limits of my world,” and tried to show that individuals understood their world as long as they could express it with their language.

He went beyond this thinking, and proposed an idea of language game. “I shall also call the whole [of language], consisting of language and the actions into which it is woven, the ‘language-game’.” (Wittgenstein, 1958, §.7). He expanded the scope to include actions. He showed the relativity in social realities by saying “The language-games are rather set up as ‘objects of comparison’ which are meant to throw light on the facts of our language by way not only of similarities, but also of dissimilarities.” (ditto, §.130). Further, he pointed out that “new types of language, new language games, as we may say, come into existence and others become obsolete and get forgotten.” (ditto, §.23). With this message he showed the changing nature of human language.

These two streams of thinking about language discarded the idea of language as the ‘mirror of nature’ (Rorty, 2009), constructed a new idea of language, and brought about an innovation in research methods of social science. Rorty (1992) named these streams as the linguistic turn.

4.3 The practice turn

In the 21st century, there is a movement that entitles the moves of recent twenty years as the practice turn (Schatzki et al, 2001). In the 20th century, knowledge about an individual and a society was separated because of methodological-individualism and methodological-collectivism in sociology. It is Giddens, Bourdieu and others who tried to conquer the dualism. Giddens (1984) shifted the problem of dualism to that of duality with structuration theory. It is through action and interaction within practices that mind, rationality and knowledge are constituted and social life is organized, reproduced and transformed (Schatzki et al, 2001, p.i).

Blumer (1986, p.71) mentioned already in the 1960s that the essence of society lies in an ongoing process of action, and that without action, any structure of relations between people is meaningless. Social systems are produced and reproduced in interaction (Giddens, 1984, p.25). Therefore, focusing not only on language but also on everyday practice became important.

At each turn from the phenomenological to the practice, epistemology of social science has expanded the scope of concern to have better understanding of social realities. It is considered significant to use these turns as a research framework to investigate the consistency of the research object, the semantic Web, to social science.
5 The semantic Web and the turns of social science

In this section we will investigate the consistency or conformance of the semantic Web to the turns of social science.

5.1 The semantic Web and the phenomenological turn

In phenomenology, there is neither an object nor an idea beforehand. Instead, it is a human who constitutes or reconstitutes a concept, and it is important that we are conscious of the constitution.

In the semantic Web, the definition of a concept is made around URI through RDFs as predicates which describe properties and relations (Fig.2).

In phenomenology, Husserl said: “the predicates are predicates of ‘something’, and this ‘something’ belongs together with the predicates, and clearly inseparably, to the nucleus in question: it is the central point of unification which we referred to above. It is the nodal point of connexion for the predicates, their ‘bearer’…” (Husserl, 2002, p.365). There detaches itself as the central noematic phase: the ‘object’, the ‘objective unity’, the ‘self-same’, the ‘determinable subject of its possible predicates’ - the pure X in abstraction from all predicates… (Husserl, 2002, pp.365-366).

URI is unique in the world, and is treated as a ‘bearer’ in the semantic Web. Here we see that the way of description of the semantic Web embodies the foundational idea of phenomenology.

In the distributed environment including the semantic Web, it is an individual person who makes the constitution of a concept. There, his/her practice conforms to OWA. It satisfies the idea mentioned by Merleau-Ponty (2002, pp.xxii-xxiii): “The phenomenological world is not the bringing to explicit expression of a pre-existing being, but the laying down of being.” It means the individual person lays down the world through describing. What is laid down is not “the world which is pre-given, taken for granted through experience” (Husserl, 1970, p.68) but “a subjective structure [Gebilde]” (ibid, p.69).
Describing is a conscious action, and can be regarded as practicing phenomenological reduction. That is, a collection of ‘subjective structure’ is the semantic Web.

As the review of prior literature shows the lack of methodology, it should be noted that how to constitute a concept remains obscure in the semantic Web. Therefore, IS research as social science is expected to contribute to the development of methodology. The methodology might be of not a third but a first person-perspective systems analysis (Kosaka, 2013).

5.2 The semantic Web and the linguistic turn

5.2.1 The semiotic turn

According to Saussure, a sign consists of signified and signifier. We articulate the world as such through a system of the signs.

As seen in Fig.2, concept constitution and naming proceed simultaneously. This simultaneous process conforms to the semiotic turn mentioned by Saussure as: “It would above all be necessary that the idea should be determinate in advance, and it is not. It would above all be necessary that the signified element should be something determined in advance, and it is not.” (Saussure, 1993, p.400). In other words, signified and signifier are made at the same time. There is no signified in advance.

In the semantic Web, constitution of concept is made around a URI, another is made around another URI and it continues further. Keeping consciousness of difference, constitution of concepts continues. If necessary, explicit description of differences can be made by using ontology such as owl.disjointClass or owl.differentFrom.

This leads to the semiotic turn by Saussure. “In the language, there are only differences, without positive terms. That is the paradoxical truth. At least, there are only differences if you are speaking either of meanings, or of signified or signifying elements.” (Saussure, 1993, p.403). Due to no positive terms, people add URIs to make differences clearer.

We should not expect an automatic creation of signs by the semantic Web. It is humans who constitute a concept. The semantic Web has only a support function that helps humans do it. It is because the semantic Web is a language first. With the use of the language, it is humans who lay down the world, i.e., the setting of the world. This conforms to the following. “Within this epistemology language is central to the process of constructing social reality - it is productive, formative and creative – it makes things happen.” (Musson et al, 2007). That is to say, describing with RDF and OWL is the use of language, and makes the world appear to us.

5.2.2 The language game

The limits of language and the language game are concerned here. According to the early Wittgenstein, language is the only means to access the reality. Since it resides between people separated by time and space, the semantic Web would not convey anything without description. It continues to be unknown because it is based on OWA. The world is laid down or made happen only when it is described in the languages.

It exists among people what can be described in the semantic Web. It is the inter-subjective version of the phrase: “the limits of our language mean the limits of our world.”

The later Wittgenstein proposed an idea of language game, and revealed that it is just of a local play. The semantic Web being locally developed is not influenced by central monitoring or control. Therefore, it is just a collection of local plays. It is possible in the semantic Web that the same thing could be named differently, and also that the same names could indicate different things. That is the reason why the semantic Web adopts non-unique name assumption. The mechanism for bridging is
prepared in ontology supported by OWL. If instances described separately happen to be the same, they can be bridged by using owl.sameAS in OWL. If so at the class level, owl.equivalentClass can be used.

To sum up, the semantic Web conforms to the idea of language game because the semantic Web adopts open-world and non-unique name assumptions.

We have to mention some technical problem in the semantic Web. Even within one person, he/she must be subject to changes of the world. Wittgenstein (1958, §23) says: “new types of language, new language games, as we may say, come into existence and others become obsolete and get forgotten.” In the semantic Web, mapping between RDFs and that between ontologies are still technical problems. They are being researched as ontology mapping, or ontology matching, for example, see Ritze et al (2010).

5.3 The semantic Web and the practice turn

After the practice turn, researchers expand their scope of attention not only to signs and languages but also to practice. That is because “the essence of society lies in an ongoing process of action” Blumer (1986, p. 71). In these days researchers try to understand social realities by paying attention to daily actions.

In the semantic Web, people are users, while they are producers, that is, prosumers. Therefore, since there is no central control person in the semantic Web, it becomes a problem how the data and knowledge produced disorderly are processed. The solution to the problem is supposed to be given by an idea of the layer cake (Fig.1). The layer cake is provided with the proof layer and the trust layer on the top. They are to treat the data in the disorderly world.

It is thought that users’ participation in the semantic Web construction is a key to gain trust. It is an idea that users are active participants to add data and knowledge to the semantic Web. In other words, users are regarded as prosumers in the semantic Web.

In contrast to the natural world, there is no truth but validity in the social world. When many people share it, validity is achieved as inter-subjectivity. This reflects an idea that firstly there is an action and that then social institution can be found within the actions. It can be seen that the layer cake of the semantic Web is intended to help a variety of people bring about realities and further reconstruct the realities. It is not just a problem of language and vocabulary. The semantic Web can work just because validity is constructed through human actions and shared. It is thought that this line of thinking conforms to the idea of the practice turn.

The use by various people realizes the proof and trust on validity as a result. It should be a fundamental idea of the semantic Web that the semantic Web takes not only description but also human actions into consideration from the beginning. Takeda (2002) says that a good ontology is not a right one but an ontology that can be shared by many people.

6 Discussions and Conclusions

There is a growing awareness of the semantic Web. It may show a different evolution from the conventional AI (Artificial Intelligence). For knowledge representation in the conventional AI, there was the domain closure assumption. The data system at hand was assumed to be complete. In other words, researchers presupposed a complete world, within which they kept a position that they could study the capabilities of representation and reasoning (Takeda, 2002). Thus, the AI was intensively researched in the area of engineering. However, practical applications of the conventional AI could hardly continue to exist due to the difficulty of maintenance.

The semantic Web is based on the open-world assumption (OWA) and the non-unique name assumption, and is a collection of knowledge and data separately constructed in various fields.
Although it is still on the way as far as technology is concerned, it can be thought that the semantic Web are being socially accepted because applications of LOD have come into existence. By contrast, the methodology, i.e., how a user develops the knowledge and data of the semantic Web, is still not clear. Research into this methodology is expected.

In order to know whether the semantic Web can be a research topic in IS research we made a theoretical investigation into the consistency of the semantic Web with the turns of social science of this one century. We have made it sure that the semantic Web has languages which are expressive enough to capture a variety of the changing social realities. It is not yet explicitly mentioned in technical terms how trust is gained in the distributed development environments. However, from the viewpoint of social science, the practice turn in particular, it can be thought that there is an unstated idea that a wide use by many people will bring validity. We can see that a social idea exceeding the technical beauty is incorporated in the semantic Web.

To our knowledge, this research is of the first one in this area. There are many problems remaining. Among them, a comparison between OWA and CWA, and an analysis of a relative significance between OWA and CWA in relation to the turns of social science in particular, and others are considered important. Further, a systems analysis methodology for prosumers is considered necessary in distributed environments such as the semantic Web. The methodology is not a traditional third person-perspective one, but one which facilitates prosumers themselves to become conscious of the conditions of enactment for their knowledge.

References


Communications of the ACM, 49(8), 15-16.

